Extra cases Setting 5

After reviewing the results in Tamvada (2023), we decided to assess performance across settings with varying numbers of covariates and different proportions of true covariates. Specifically, this involved using an intermediate sample size ($p = 500$) and changing the number of true covariates. While the previous analysis kept the number of covariates constant, in the new setup we maintained the same ratio of variables to ensure consistency.

1 Simulations

Each simulation was run 20 times with a sample size of $n = 400$. The covariate data, X, for the p predictors were generated from a multivariate normal distribution with a mean of zero, $X \sim \mathcal{N}(0, \Sigma)$. The covariance matrix Σ was defined to have an exchangeable correlation structure where the correlation between the true influential covariates was $\rho = 0.5$. The simulations were designed for a non-proportional hazards setting. For a scenario with p total predictors and k true predictors (e.g., $k = 20$), the coefficient vectors for the two events (β_1 and β_2) were defined as:

$$\beta_1 = (\underbrace{1, \ldots, 1}_{k/2}, \underbrace{0, \ldots, 0}_{p-k/2})^T$$

$$\beta_2 = (\underbrace{-1, \ldots, -1}_{k/2}, \underbrace{0, \ldots, 0}_{p-k/2})^T$$

In this structure, $k/2$ predictors influence event 1, and a distinct set of $k/2$ predictors influences the competing event, with all other $p - k$ predictors having no effect.

The competing models for the experiment are the following:

- **Elastic-net casebase (enet-casebase):** The penalty parameter, λ_{min}, is tuned using 5-fold cross-validation due to computational constraints. To fit a Weibull hazard, the time variable is included as a covariate in the form $\log(t)$. This model is trained with SVRG.
- **Elastic-net casebase (enet-casebase-Acc):** This model is also trained with AccSVRG.
- **Elastic-net Independent Cox (enet-iCR):** For this model, the penalty parameter, λ_{min}, is tuned using 10-fold cross-validation, with the partial likelihood deviance used as the performance metric.
- **Elastic-net Cox with shared penalty (enet-penCR):** This model's penalty parameter, λ_{min}, is tuned using 10-fold cross-validation across a 30 ×30 grid. The Brier score for event 1 serves as the performance metric.

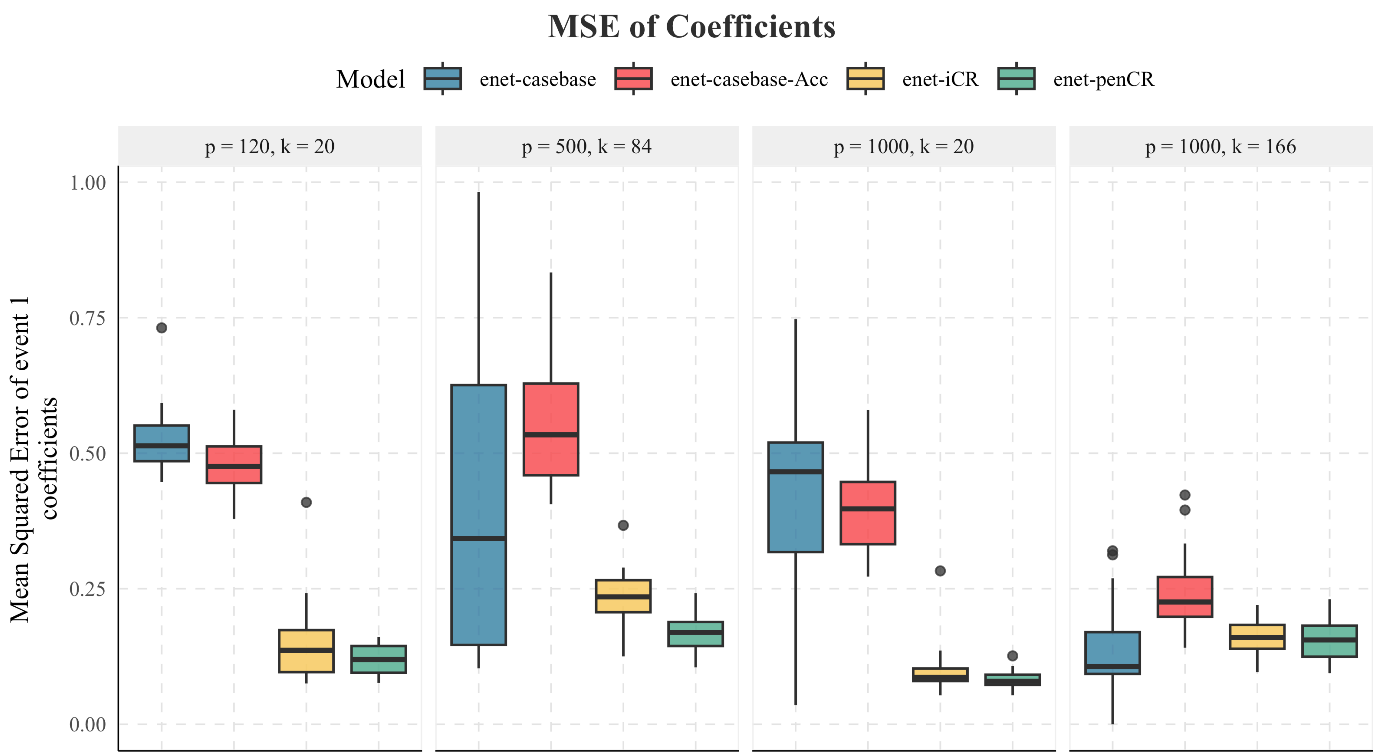
1.1 Results

The following plots show the MSE of event 1 coefficients, along with variable selection metrics such as Sensitivity, Specificity, and MCC. As observed in previous analyses, the casebase models produce coefficients with higher squared error. However, it appears that the Acc method tends to have higher error levels, especially in scenarios where the proportion of true variables is preserved.

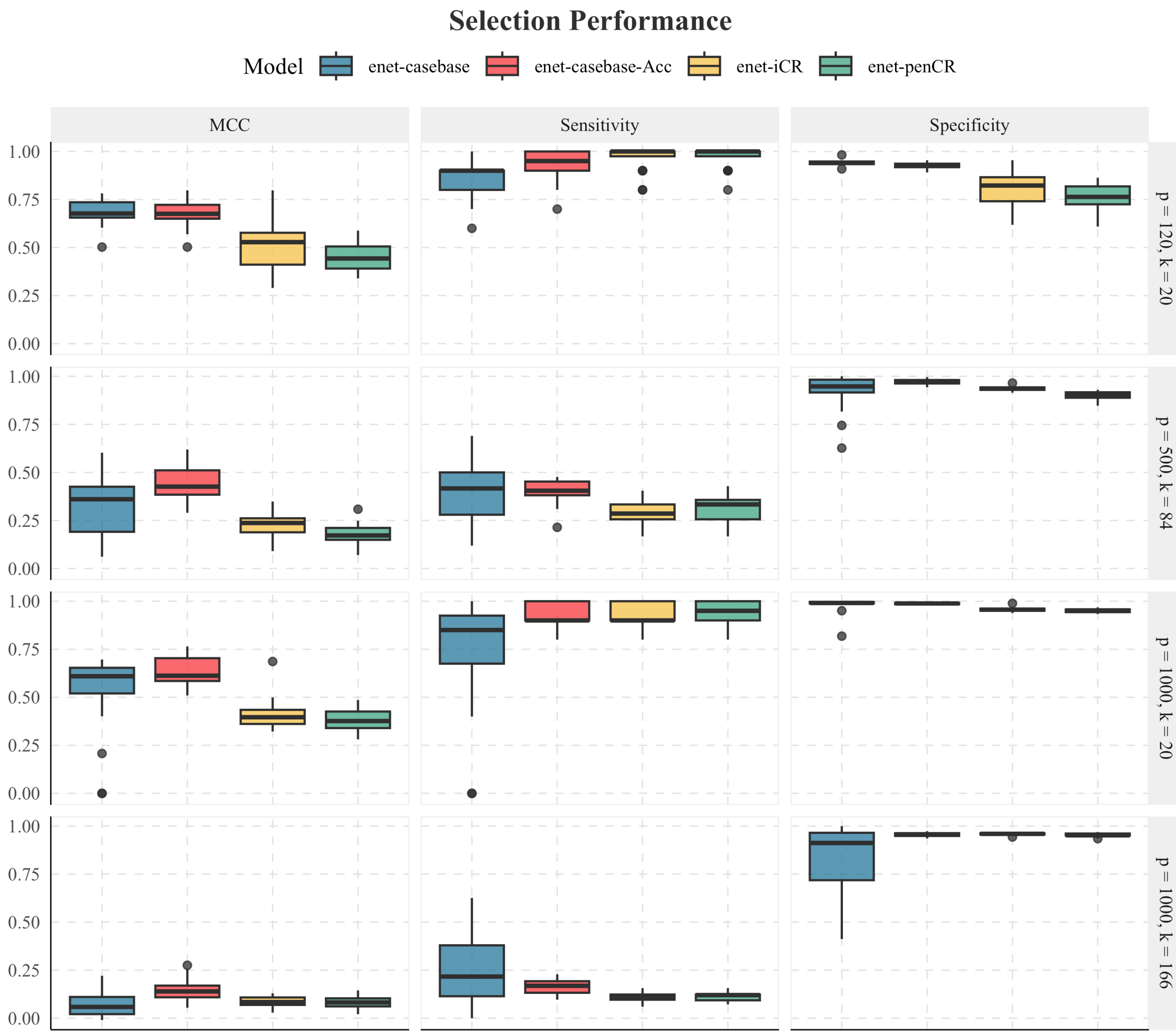
The second plot illustrates the variable selection performance. In this case, the proposed methods consistently achieve a higher MCC across all scenarios. Notably, the proposed model trained using the Acc method exhibits slightly better performance than the one trained with the SVRG method, except for Sensitivity in the case where $p = 1000$ and $k = 166$.

```
coefs_mod %>%
  group_by(name, model, setting, p, k) %>%
  summarise(bias_mse = vec_mse_bias(coef, p, k)) %>%
  arrange(p, k) %>%
  filter(!str_detect(model, "post")) %>%
  ggplot(aes(x = model, y = bias_mse,
             fill = model)) +
  geom_boxplot() +
  facet_grid(~fct_inorder(setting)) +
   theme(axis.text.x = #element_text(angle = 90, vjust = 0.5, hjust=1)
          element_blank()) +
  labs(title = "MSE of Coefficients",
     fill = "Model",
     y = "Mean Squared Error of event 1\ncoefficients",
     x = element_blank())
```

```
#> Warning: Removed 3 rows containing non-finite outside the scale range
#> (`stat_boxplot()`).
```




```
coefs_mod %>%
  group_by(name, model, setting, p, k) %>%
  summarise(selec_mea = vec_varsel_perc(coef, p, k)) %>%
  unnest(selec_mea) %>%
  arrange(p, k) %>%
  filter(!str_detect(model, "post")) %>%
  pivot_longer(Sensitivity:MCC,
          names_to = "metric",
          values_to = "value") %>%
  ggplot(aes(x = model, y = value,
         fill = model)) +
  geom_boxplot(show.legend = T) +
  facet_grid(fct_inorder(setting)~metric) +
  theme(axis.text.x = #element_text(angle = 90, vjust = 0.5, hjust=1)
        element_blank()) +
  labs(title = "Selection Performance",
     fill = "Model",
     y = element_blank(),
     x = element_blank())
```



References

Tamvada, Nirupama. 2023. "Penalized Competing Risks Analysis Using Casebase Sampling." University of British Columbia. https://doi.org/10.14288/1.0435526.